PREDICTION OF PM_{10} CONCENTRATIONS USING EXTREME VALUE DISTRIBUTIONS (EVD) : CLASSICAL AND BAYESIAN APPROACHES

HASFAZILAH BINTI AHMAT

UNIVERSITI SAINS MALAYSIA
2016

PREDICTION OF PM_{10} CONCENTRATIONS USING EXTREME VALUE DISTRIBUTIONS (EVD) : CLASSICAL AND BAYESIAN APPROACHES

by

HASFAZILAH BINTI AHMAT

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2016

ACKNOWLEDGEMENT

In the Name of Allah, the Most Beneficent and the Most Merciful. All the praises and thanks be to Allah, the Lord of the *worlds* who has showed us the straight way, the way of those on whom He have bestowed His Grace, not (the way) of those who earned His Anger, nor of those who went astray.

The writing of this dissertation has been one of the most significant academic challenges I have ever had to face. Without the support, patience and guidance of a few significant persons in my life, this study would not have been completed.

First and foremost, I would like to infinitely thank the Ministry of Higher Education, Malaysia and Universiti Teknologi MARA for the scholarship and study leave granted. Not forgetting the Department of Environment, Malaysia for the invaluable information (monitoring records) to be used in this study.

It is difficult to overstate my gratitude and appreciation to my supervisor, Assoc. Prof Ahmad Shukri bin Yahaya. Throughout this study - with his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, his wisdom and encouragement - I would have been lost without him. In addition, this appreciation goes to the expert in the field of environment, Prof Dr. Nor Azam Ramli, my co-supervisor, for his support, advice, guidance and constructive comments throughout the study period.

I owe my deepest gratitude to Mohd Fazal, my husband, without whom this effort would have been worth nothing. Your love, support and constant patience have taught me so much about sacrifice and compromise. The same goes to my lovely daughters, Farzana and Farwiizah, who bear with their student mother.

My heartiest gratitude to my mother, who have always supported, encouraged and believed in me, in all my endeavours. My sincere gratitude to my family and in-laws for all their encouragement and profound understanding. To my beloved late father and father in-law for their previous supports.

Special thank goes to my friends and colleagues who inspired my final effort. I wish to thank my friends in the Clean Air Research Group (CARE) and at the Postgraduate room for helping me get through the difficult times, and for all the emotional support and caring they provided during the postgraduate years.

Not to forget, special thanks also to the Facebook *Doctorate Support Group*'s and *MS Word for Thesis*'s admin and members for the knowledge sharing and constant words of encouragement. I wish to thank all staff at the School of Civil Engineering for providing a supportive environment for me.

Last but not least, sincere appreciation to my entire extended family who for providing a loving environment and particularly supportive. To them I dedicate this thesis.

TABLE OF CONTENT

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AQG	Air Quality Guidelines
ASMA	Alam Sekitar Malaysia Sdn. Bhd.
BUGS	Bayesian Inference using Gibbs Sampling
CDF	Cummulative Distribution Function
DOE	Department of Environment
EVD	Extreme Value Distribution
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution
IA	Index of Accuracy
MAAQG	Malaysia Ambient Air Quality Guidelines
MAE	Mean Absolute Error
MCMC	Markov Chain Monte Carlo
MIS	Method of Independent Storm
MK	Mann-Kendall's
MLE	Maximum Likelihood Estimator

MOM	Method of Moments
NAE	Normalised Absolute Error
PA	Prediction Accuracy
PDF	Probability Density Function
PM	Particulate Matter
PM_{10}	Particulate Matter with aerodynamic diameter of less than 10μm
POT	Peak Over Threshold
PWM	Probability-Weighted Moments
R^2	Coefficient of Determination
RMSE	Root Mean Squared Error
SPSS	Statistical Package for the Social Sciences
WHO	World Health Organisation

PERAMALAN KEPEKATAN PM_{10} MENGGUNAKAN TABURAN NILAI MELAMPAU (EVD) : PENDEKATAN KLASIK DAN BAYESIAN

ABSTRAK

Keadaan zarahan tinggi (jerebu) secara umumnya dikaitkan dengan kehadiran PM_{10} atau $PM_{2.5}$. Ia adalah penting untuk memaklumkan kepada umum terhadap tahap PM_{10} dan kepentingannya supaya langkah-langkah penyesuaian yang lebih berkesan dapat diambil bagi kalangan umum yang terjejas. Kajian ini dijalankan dengan objektif untuk membandingkan Taburan Nilai Melampau (EVD) menggunakan pendekatan konvensional dan Bayesian dan menggunakan taburan terbaik untuk peramalan kepekatan PM_{10} pada masa hadapan. Ketika ini, tiada pendekatan Bayesian di dalam kajian kepekatan PM_{10}. Rekod daripada lapan stesen pengawasan di Semenanjung Malaysia telah dipilih untuk tempoh 1 Januari 2000 hingga 31 Disember 2012 selepas analisis awal untuk menilai kewujudan nilai melampau. Taburan dengan pengukuran ralat yang terkecil dan pengukuran kejituan tertinggi di lima stesen pemantauan – Bukit Rambai, Jerantut, Nilai, Pasir Gudang dan Shah Alam adalah taburan menggunakan kaedah Bayesian dengan kebolehjadian GEV dan taburan prior tanpa maklumat menggunakan taburan seragam. Walau bagaimanapun, bagi Klang dan Seberang Jaya taburan EVD GEV disimpulkan sebagai taburan yang terbaik dan EVD dua parameter Weibull adalah taburan terbaik untuk Perai. Pendekatan Bayesian adalah lebih unggul dari kaedah konvensional apabila menggunakan data maksimum harian dan boleh digunakan untuk menilai tahap kepekatan tinggi PM_{10} untuk penggubal dasar melaksanakan dasar-dasar yang lebih berkesan untuk mewujudkan persekitaran yang lebih bersih.

PREDICTION OF PM_{10} CONCENTRATIONS USING EXTREME VALUE DISTRIBUTIONS (EVD) : CLASSICAL AND BAYESIAN APPROACHES

ABSTRACT

High particulate event (haze) is generally associated with presence of PM_{10} or $PM_{2.5}$. It is important to make known to public of PM_{10} level and its importance for more effective adaptation measures among the affected public. This study was conducted with the objectives to compare the best Extreme Value Distributions (EVD) using the conventional and Bayesian approaches and use the best distribution for the prediction of future PM_{10} exceedances. Currently, there is none on the application of Bayesian approach in the study of PM_{10} concentrations. Records from eight monitoring stations in the Peninsular Malaysia were selected for the period of 1st January 2000 to 31st December 2012 after preliminary analysis to check for the existance of extreme values. The distribution with the smallest error measures and highest accuracy measures in five of the monitoring stations – Bukit Rambai, Jerantut, Nilai, Pasir Gudang and Shah Alam was the Bayesian GEV likelihood with uniform non-informative prior distribution. However, for Klang and Seberang Jaya the EVD GEV distribution was concluded as the best distribution and EVD two-parameter Weibull was the best distribution for Perai. The Bayesian approach is superior than the conventional method using the daily maximum data and can be used to assess high level of PM_{10} concentrations for the policy makers to implement effective policies to create cleaner environment.

CHAPTER 1
INTRODUCTION

1.1 AIR POLLUTION

Air pollution is the global issue that has been one of major concerns in developed and developing countries. Exposure to air pollution is mainly outside the control of individuals (Ramli *et al.*, 2003). Air pollution is defined as the contamination of the atmosphere by foreign substances in the form of gaseous, liquid or fine particles suspended in air in large numbers which provide harmful effects to plants, animals, humans and human property or the global environment (Jamal *et al.*, 2004; Department of Environment Malaysia, 2013). Kabir *et al.* (2012) stated that air pollution, both indoor and outdoor, is often considered as the major cause of environmental health problems. Indoor air pollution is identified as a major contributor to mortality and morbidity from acute lower respiratory illness in children (Franklin, 2007).

The cause of air pollution can be from the primary sources or secondary sources. Primary sources are the direct result of the process to produce pollutants in air for example the emission of sulphur dioxide from the factories and vehicles combustions. On the other hand, secondary sources are the ones that are formed within the atmosphere due to reactions of primary pollutants mostly from the chemical oxidation of atmospheric gasses. Smog is amongst the secondary pollutants which is originated from vehicular emission from internal combustion engines and industrial fumes that react in the atmosphere with sunlight to form

photochemical smog (Department of Environment, 1996; Environment Protection Authority, 2004).

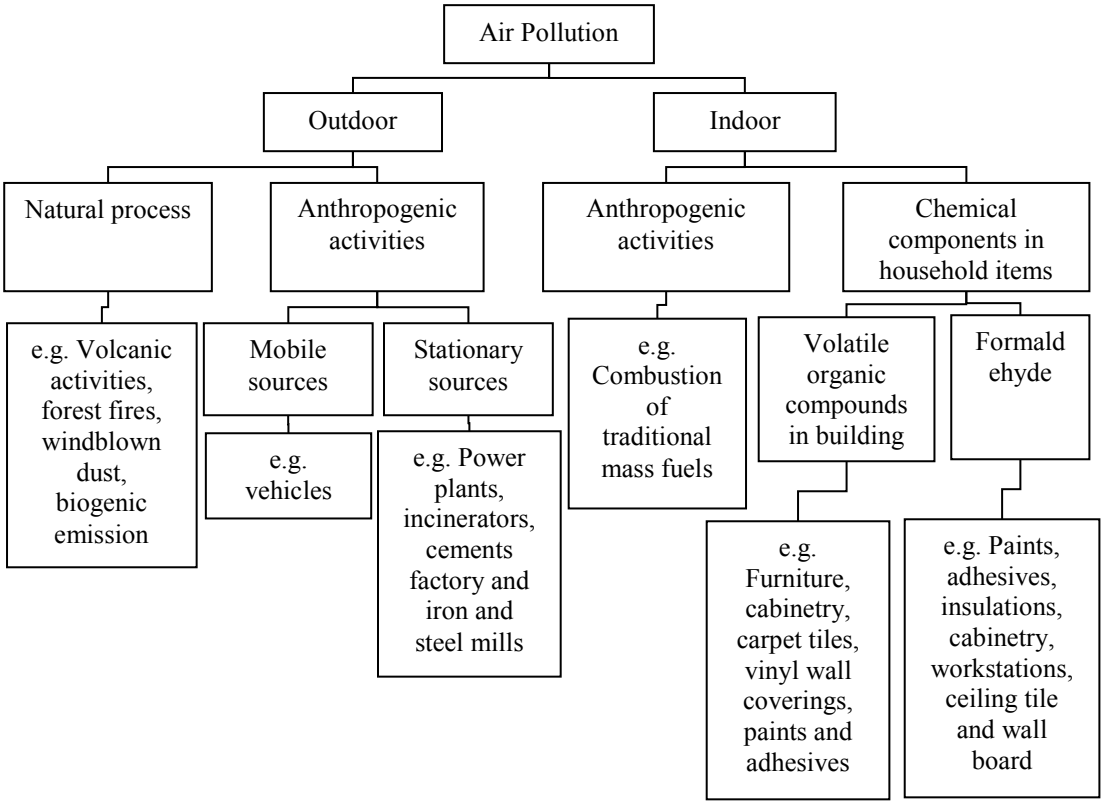

Figure 1.1 Summary of the category of air pollution and its sources

As illustrated in Figure 1.1, outdoor air pollution occurs as a result of natural processes (volcanic activity, forest fires, windblown dust and biogenic emissions from vegetation) coupled with anthropogenic activities. Motor vehicles either public or private vehicles which everyone rely on them to fulfill daily basic needs of transportation are the major contributors for mobile sources. The incomplete combustion of the vehicles emit carbon monoxide, the major pollutant together with nitrogen oxides which are produced both via naturally and man-made process. The industrial processes for instance power plants, incinerators, cements factory and iron and steel mills as stationary sources, release abundance of hydrocarbons, organic compounds and chemicals into the air, are among the anthropogenic activities that

contributed to air pollution (Department of Environment, 1996; Ramli *et al.*, 2003; Jamal *et al.*, 2004).

The indoor air pollution occurs mainly from the combustion of traditional biomass fuels (wood, cow dung, and crop wastes) which is predominant for poor populations in many developing countries (Khalequzzaman *et al.*, 2007; Begum *et al.*, 2009). This combustion is responsible for many indoor pollutants, including carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), volatile organic compounds (VOCs), and particulate matter (Khalequzzaman *et al.*, 2011). Kabir *et al.* (2012) indicated that the composition and concentrations of the different components in indoor air vary widely and are influenced by human activities. A recent indoor study which involved 628 United Arab Emirates (UAE) personal residences indicated that air conditioning, smoking, and attached kitchens were significantly associated with levels of carbon monoxide. In addition, vehicles parked within five meters from home, central air conditioning, and having attached kitchens also contributed to the increase level of indoor concentrations of PM_{10} (Funk *et al.*, 2014) which is in line with the results of Kabir *et al.* (2012).

Depending on the place one lives, one source may be more relevant to oneself than another. In urban areas, transport is a major contributor in ozone and particle pollution. On the other hand, in rural areas, agricultural burning, coal heaters and solid fuel heaters are the main sources of particles and ozone is not really a big issue. Air pollution comes from seasonal changes throughout the year. Particle pollution from bushfires is more common in summer and pollution from wood heaters is most likely in winter (Department of Environment Climate Change and Water NSW, 2010).